DOCUMENT RESUME

ED 268 246 CE 043 931

AUTHOR	Carlisle, Ysanne
TITLE	Easy Statistics. Supervising: Technical Aspects of Supervision. The Choice Series #34. A Self Learning Opportunity.
INSTITUTION	Northern Regional Management Centre, Washington (England).
SPONS AGENCY	Manpower Services Commission, London (England).
REPORT NO	ISBN-0-948023-21-X
PUB DATE	85
NOTE	68p.; For related documents, see ED 266 310-312 and CE 043 921-948.
AVAILABLE FROM	Northern Regional Management Centre, Derwent House, Washington Town Centre, Washington, Tyne & Wear NE38 7ST, England ($15.00 plus postage; 15% discount on 50, 25% on 100; includes audiotape).
PUB TYPE	Guides - Classroom Use - Materials (For Learner) (051)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Adult Education; Behavioral Objectives; *Decision Making; Learning Activities; Management Development; Self Evaluation (Individuals); *Statistical Analysis; Statistical Data; *Statistics; Student Evaluation; *Supervision; Supervisor Qualifications; *Supervisory Methods; *Supervisory Training
IDENTIFIERS	Great Britain

ABSTRACT

This learning unit on easy statistics is one in the Choice Series, a self-learning development program for supervisors. Purpose stated for the approximately eight-hour-long unit is to enable the supervisor to use statistics to improve his/her decision-making ability, interpret statistics and form judgments on the interpretations other people make, assess statistically a work team's performance, and use quick and practical statistical tests at work. An introduction provides an overview of the unit and lists unit objectives. Parts A through D contain informative material, activities and answer keys, self-checks and answer keys, and a summary. Topics include describing data, some basic statistics, a basic statistical test, and relationships and trends. Part E contains performance checks--an end check to be completed alone and a tutor check and work-based project. Part F consists of a brief unit review, extensions (print sources of additional information, such as units or sections in books), and a list of references. (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

An Open Tech Project

SUPERVISING

TECHNICAL ASPECTS OF SUPERVISION

34: Easy Statistics

ED268246

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

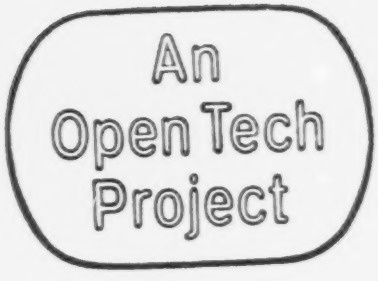
"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

H Pardue

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

34: EASY STATISTICS

Ysanne Carlisle



The CHOICE programme is an Open Tech sponsored project managed by the Northern Regional Management Centre in conjunction with the Northern Consortium of Colleges and the National Examinations Board for Supervisory Studies.

© Crown Copyright 1985 Published by permission of the Controller of Her Majesty's Stationery Office.

ISBN 0 948023 21 X

First published in the U K - 1985

Northern Regional Management Centre
Derwent House, Washington Town Centre
Washington, Tyne & Wear NE38 7ST

This Work was produced under an Open Tech contract with the Manpower Services Commission. The views expressed are those of the authors and do not necessarily reflect those of the MSC or any Government Department.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission in writing from the publisher.

Ysanne Carlisle
Easy Statistics (Number 34 in the CHOICE SERIES of learning units)

Series Editor: Harry V Pardue
Northern Regional Management Centre.

CONTENTS PAGE

PART A DESCRIBING DATA

1 Introduction 2

2 Averages 2

3 Finding the mean from a table 4

4 The mean of continuous data 6

5 Range of data 8

6 The standard deviation 12

7 Using ranges and averages 12

8 Summary 14

PART B SOME BASIC STATISTICS

1 Introduction 15

2 Spotting rogue values 15

3 The square root 18

4 The standard deviation 19

5 An acceptable level for testing 21

6 Summary 25

PART C OUR BASIC STATISTICAL TEST

1 Introduction 26

2 Our statistical test 26

3 Comparing two sets of figures 31

4 Sample size 35

5 Summary 35

PART D RELATIONSHIPS AND TRENDS

1 Introduction 36

2 Relationships in data 36

3 Spotting a trend 41

4 Summary 46

PART E PERFORMANCE CHECKS

1 End Check 47

2 Tutor Check 52

3 Work based project 54

PART F UNIT REVIEW

1 Return to objectives 55

2 Extensions 56

3 References 60

WORKING GUIDE TO THIS UNIT IN YOUR DEVELOPMENT PROGRAMME

This Unit is now your personal property - it's not for handing back to your training officer or to a tutor. It is your working book.

The Contents will increase in value to you each time you write on one of its pages. This is a working text for your personal notes - without notes its value is limited.

Qualifications forget about certificates and diplomas - you don't need them for this programme. What you do need is the wish:-

* to do your job better
* for personal development

Your study time is really up to you - we guess you will need about 8 hours in total. But it's your time - spend as long as you like.

How am I doing? you can judge for yourself by checking out how well you cope with the tasks we call Activities and Self Checks which occur throughout the unit. You can also get the Tutor Check and/or the Work Based Project to your tutor for an unbiased view of your progress.

Work Pace it's really up to you and depends on your interest. However, if you do well on the Activities and Self Checks, you could afford to speed up a little. Of course, if you don't do as well as you expect then you could slow down a little or flick back into the unit for a quick review of the more difficult ideas.

When in doubt - ask or seek help from:

* your family and friends
* colleagues at work
* your company training staff
* other open learners
* your tutor, by phone, by a visit to your support centre or by dropping your tutor a line through the post.

Don't worry if what you learn doesn't seem to be of use to you now, remember that times change - it may be useful one day. Anyway, just knowing more improves personal confidence.

So now make a start:-

34: EASY STATISTICS by Ysanne Carlisle

ABOUT THIS UNIT

Statistics is a subject which can help us to understand numerical information, but it can also be used as a tool to help us in our decision making. We can use statistical techniques to help us make predictions and comparisons, spot trends, determine whether say, increased output leads to more scrap. To use statistics as a tool in this way, we need not know something of the theory behind the techniques. The aim of this unit is to provide you with a battery of easy statistical techniques to use in practical situations.

IN THIS UNIT WE WILL:

* Develop an ability to interpret statistics from a sound base
* Explore the predictive potential of statistics
* See how statistics can help decision taking
* Build up a battery of quick practical tests.

Audio tape Side 1

Before you continue through the unit, you should now listen to Side 1 of the audio tape that accompanies the unit. You will find it useful if you refer to the unit objectives shown below whilst you listen to the tape. You may wish to make the objectives clearer to yourself by making a note or two on the page as you listen.

UNIT OBJECTIVES

On completion of this unit you will be BETTER ABLE to:

* **Use statistics to improve your decision-making ability**
* **Interpret statistics and form judgements on the interpretations other people make**
* **Assess statistically, the performance of your work team**
* **Use quick and practical statistical tests at work**

PART A DESCRIBING DATA

1 Introduction

REFERENCE 1
Page 60

Betts points out the need for supervisors to have technical competence. Competence in some basic statistical techniques can be very useful for evaluating figures at work.

Two useful measures which can be used to help us make sense of a set of numbers are the average and a measure of spread. An average is a figure which is used to represent that set of figures. A measure of spread tells us how widely different our individual figures are. In this part of the unit we'll look at the average and measures of spread.

We'll also look at a way of modifying the average because in certain circumstances this can be useful.

2 Averages

The average is important because it uses every value in a sample and is useful for other statistical tests.

It can be calculated with certainty for any set of figures.

If we know all the values in a sample it can be found by dividing the total of our values by the number of values.

Here is an example. We have 10 values:

2.1; 3.1; 6.2; 3.1; 3.2; 4.0; 2.9; 1.3; 1.1; 1.0:

The total of these values is 28.

The total divided by the number of values is our mean. So 28 divided by 10 gives 2.8 as our mean.

SELF CHECK 1
Time guide
3 mins

Find the mean of the following values.

30; 52; 44; 40; 32; 22; 32; 28

Total of values = * _____

Number of values = * _____

Therefore mean = * _____

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. The total of our values is 280. The number of values is 8. The mean is found by dividing the total by the number of values, and is 35.

While the mean is a useful and important statistic it can be affected by an extreme value, for example suppose we had recorded our first figure in our previous example wrongly. Let's assume we misheard and it was recorded as 13 instead of 30. Our figures would then be:

13; 52; 44; 40; 32; 22; 32; 28:

SELF CHECK 2
Time Guide
2 mins

What is the new mean?

* _____

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. Our new mean is very nearly 33. Our new total is 263. Our number of values is still 8. Our previous mean was 35.

There is quite a difference between these means of 35 and 33.

Perhaps we should ask ourselves whether the extreme figure in our example - 13 - is genuine or is a mistake. If it is genuine then is there an alternative statistic to the mean? Before we answer this last question look at this activity:

ACTIVITY 1
Time guide
2 mins

Respond to the questions in the space provided.

You have been given the scrap rates for each of your work team. One of these values looks far too large!

1. Why might the mean not be a good average to use to represent the group?

   * _____

   _____

2. What other question would you ask yourself about this large value?

   * _____

   _____

---

1 Because we know that an extreme value can seriously affect the mean - we need to be a little cautious.

2 We could check out to make sure that the extreme value is not a mistake.

Because it is easily affected by extreme values, we may not always want to use the mean as a measure to represent a group of values. Let's suppose we checked all our values very carefully and were sure that all of them had been properly recorded. There are other measures we could find. The two most common are -

THE MODE which we can define as the most frequently occurring value

EXTENSION 1
Page 56

and

THE MEDIAN which we can define as the middle value in a set of figures.

In this unit we'll concentrate on the mean, because it has more practical uses.

3 Finding the Mean From a Table

We have already seen how to find the mean of a set of figures. When there's only a few figures, then our method is quite quick. Often, however, there will be a lot of figures which can be tedious to handle; adding up, say, more than 30 figures can be a bore. Mistakes are also easy to make, so we make life easier by using a table to contain our data. In this case it's a special kind of table used regularly in statistics - it's called a FREQUENCY DISTRIBUTION TABLE. We'll soon see why and see how helpful it is.

Let's see how to take advantage of this kind of table: We could record, in one column, all the numbers of defects produced in an office. Let's assume that in a particular day no-one produced less than one defect and no one produced more than 5. We can record the values 1 to 5 and say how many operators produced 1 defect, how many produced 2 defects and so on:

No. of Defects	No. of Operators
1	12
2	11
3	20
4	19
5	8

This is better than listing the defects produced by each of 70 operators. x is a label we give to our values as a shorthand way of describing them. f stands for the frequency - the number of times each x value occurs. So our table would be like this:

x	f
1	12
2	11
3	20
4	19
5	8

To find the mean from this table we must carry out the following steps.

Step 1 - Multiply the values by their frequencies - f X x

Step 2 - Add up the total of the values multiplied by their frequencies to find the total of our values - sum of all 'f X x'

Step 3 - Add up the frequencies to find the total number of values. - sum of all 'f'

Step 4 - Divide the total of our values as found in Step 2, by the total number of values, as found in Step 3. - sum of all 'f X x' divided by sum of all 'f'

So we can use this table to calculate the mean of our defects:

SELF CHECK 3
Time guide
5 mins

I've put the data in. Complete the table and calculate the mean, step by step, as we've described:

x	f	f X x
1	12	12
2	11	22
3	20	60
4	19	76
5	8	40
Totals	* ____	* ____

Mean = * _____ divided by * _____ = * _____

5

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. Your completed table should look like the one below:

x	f	f X x
1	12	12
2	11	22
3	20	60
4	19	76
5	8	40
Totals	70	210

The mean is 210 divided by 70 = 3.

Another more common way of writing "f X x" is "fx". There is also a shorthand way of writing "the sum of"; it is Σ - the Greek symbol sigma.

So using these "shorthands" we can write our mean as:

Σfx divided by Σf or $\frac{\Sigma fx}{\Sigma f}$

This seems easier to remember than our four steps we noted earlier.

4 The Mean of Continuous Data

In our last example we calculated the mean of defects produced by a group of operators. But some data doesn't come like this. Let's see what continuous data is through an example.

A large wine and spirit merchant was interested in the average strength of the drink it sold. The strength is measured by a figure called "proof" and is given as a percentage - we see this figure on most bottles we buy.

Now the strength of each drink could be anywhere between say 6.7% to 87%. Now let's say that every bottle had a different proof strength. We'd have hundreds of x values - so our frequency distribution table would be far too large to handle. So what can we do to help us handle this kind of data?

The simple answer is to "group" the data eg we could count up the sales of drink whose strength lies between 5% and 15%; and that between 15% to 25%; and so on.

Let's move on to an example with actual data of alchohol strength and sales:

Groups % Proof	No of gallons sold
5 - 15	6
15 - 25	4
25 - 35	8
35 - 45	5
45 - 55	6
55 - 65	5
65 - 75	3
75 - 85	2
85 - 95	1

It doesn't matter here how we group our data - as long as we define them - each group should be the same size. In this example a group covers a range of 10%.
We also decide that our group of 5 - 15 means from 5 to just under 15 and our group 15 to 25 means from 15 to just under 25 and so on continuously to 95.

The calculations of our mean is done just as before but we need one more column. This is the average proof of each group - we therefore note the mid-point of each "proof" group.

SELF CHECK 4
Time guide
8 mins

Here's the data gathered by our wine and spirit merchant - complete the table and calculate the mean.

Group % Proof	% Proof middle-point x	Gallons Sold f	fx
5 - 15	10	6	*
15 - 25	20	4	*
25 - 35	*	8	*
35 - 45	*	5	*
45 - 55	*	6	*
55 - 65	*	5	*
65 - 75	*	3	*
75 - 85	*	2	*
85 - 95	*	1	*
	Totals	*	*

Mean = *

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. The total of f is 40, total fx is 1640, the mean is total fx divided by total f and is 1640 divided by 40 to give 41. Here is the completed table as a check.

Group % Proof	% Proof middle-point x	Gallons Sold f	fx
5 - 15	10	6	60
15 - 25	20	4	80
25 - 35	30	8	240
35 - 45	40	5	200
45 - 55	50	6	300
55 - 65	60	5	300
65 - 75	70	3	210
75 - 85	80	2	160
85 - 95	90	1	90

5 Range of Data

The most straightforward way of looking at how widely spread a set of figures is, is to look at the range. This is a word we have already used. The range is

THE DIFFERENCE BETWEEN THE HIGHEST AND LOWEST VALUE.

Let's look at an example.

Here is a set of figures showing the average number of times people from each of 8 groups visited a restaurant in a year.

6.4; 7.8; 8.1; 8.2; 8.4; 8.7; 9.0; 9.1

The highest value is 9.1.

The lowest value is 6.4.

The differences between the two is 9.1 - 6.4 which is 2.7.

THE LARGER THE RANGE THE MORE WIDELY SPREAD OUR DATA IS.

SELF CHECK 5
Time guide
2 mins

Suppose these earlier figures were for restaurant users in the Winter months. Here's the data for the Summer months in the same restaurant - what is the range here:

5.9; 6.2; 6.4; 6.5; 7.2; 8.1; 9.0; 9.1

The highest value is _____.

The lowest value is _____.

The difference between the two is the range which is _____.

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. The highest value in our set of figures is 9.1. The lowest value is 5.9. We find the range by taking the lowest value from the highest value, in other words by finding the difference between the two. The difference between the two is 9.1 - 5.9 or 3.2 which is our range.

Let's compare the two sets of figures we've looked at so far to see which is the more widely spread.

SELF CHECK 6
Time guide
2 mins

In our example the "Winter" range was 2.7 and the"Summer" range was 3.2. Write TRUE or FALSE in the space provided beside the sentence below.

The data for the "Summer" is more widely spread than the "Winter" data.

_____

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. The most correct answer is True. 3.2 was the summer range - this is larger than 2.7 which was the winter range. The larger the range, the more spread is our data.

However as a measure of how widely spread our data is, the range is not always the best guide. Let's see if we can spot the problem with the range from the next set of figures.

Here is a set of figures showing the average number of times the people in each of eight different age groups visited a recreation centre in a month.

1.4; 6.3; 7.1; 12.4; 7.2; 7.4; 7.5; 7.9

There are eight figures in this set. The highest and lowest figures seem to be somehow different to the other six figures. Now the Range, as we have defined it, will take all the data into account and is the difference between the highest and lowest figures.

We'll look at this example more closely.

The range is 12.4 - 1.4 = 11

If we ignore the two extreme values 12.4 and 1.4.

The range will be 7.9 - 6.3 = 1.6.

1.6 looks a better measure of the spread than 11.

Although we have missed out two of our figures to work out this range, the remaining 6 are very closely grouped. We can miss out extreme values to give us a better picture of spread when this is sensible. This would be especially true if we had a lot of figures where one or two "wild" or extreme figures were present.

SO WE CAN USE OUR JUDGEMENT WHEN GIVING THE RANGE

What we really want to do is to give a figure for the range which best represents the data as a whole.

Let's review our work on the spread of data with a self check.

SELF CHECK 7
Time guide
5 mins

Respond to each of the following statements by writing either TRUE or FALSE in the space provided.

1 The range is the difference between the highest and lowest value in a set of figures. _____

2 Here is a set of figures -

2.4; 2.5; 2.7; 2.9; 3.0

The range of these figures is 0.5. _____

3 Here is another set of figures

2.1; 2.5; 2.7; 2.9; 3.7

The range of these figures is 1.6. _____

4 The range can be affected by extreme values. _____

5 We can sometimes get a better picture of the spread of data by missing out extreme values before we calculate a figure for the range. _____

6 The set of figures in statement 2 is more spread out than the set of figures in statement 3. _____

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully 1) this defines the range, so the correct response would be true 2) this statement gave 5 figures and stated that the range was 0.5, this is false. The range is the difference between the highest and lowest values. The highest value was 3.0 and the lowest value 2.4. The range is therefore 0.6. 3) the range is calculated correctly so the correct response is true 4) This is true, the range can be affected by extreme values. 5) is true, sometimes it is better to leave out extreme values to get a better picture of spread. 6) this statement is false because the larger the range the more spread out the data.

6 The Standard Deviation

An alternative measure of spread to the range is the standard deviation. It is a bit more difficult to work out than the range, but a lot more useful, as we'll see later. The standard deviation is a measure of howwidely spread our data is around the mean. This extension gives us the chance to look at the standard deviation in detail, however all we need to know here is that the standard deviation is an alternative measure to the range and it:

EXTENSION 2
page 56

MEASURES HOW SPREAD OUR DATA IS FROM THE MEAN.

As in the case of the range, the larger the standard deviation the more widely spread is the data.

SELF CHECK 8
Time guide
1 mins

Before moving on to the next section, let's summarise our understanding of the standard deviation by doing the following self check.

Respond to each statement by filling in the blanks.

1 The standard deviation measures how widely spread our data is from _____.

2 One set of figures has a standard deviation of 0.2. A second set of figures has a standard deviation of 0.35. The first set of figures is _____ spread out than the second.

The answer to this self check is here in the following words, so you'll need to check out your answers carefully 1) the standard deviation measures how widely spread our data is from the mean 2) The first set of figures has a smaller standard deviation, which means that it is less spread out than the second.

7 Using Ranges and Averages

As we've already said, the average is a measure which can be used to describe data. It gives us a measure which can be used to represent all our figures. The range tells us how spread out our figures are. So, between them, the average and the range can be used to give us a fairly good picture of a set of figures.

Let's take a look at two sets of output figures. Two production lines, starting a new product run, each produced the following number of good items over a period of a week. Each figure is a daily output figure.

Line A	150	225	400	450	500
Line B	90	235	350	490	560

SELF CHECK 9
Time guide
5 mins

Let's remind ourselves of how to find the mean by filling in the blanks below to find the mean weekly output for lines A and B.

1 Line A

Total number of items produced in the week _____.

Total number of recorded figures _____.

Mean output is _____ divided by _____.

Which is _____ items.

2 Line B

Total number of items produced in the week _____.

Total number of recorded figures _____.

Mean output is _____ divided by _____.

Which is _____ items.

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. We'll look at each line separately. 1) Line A, The total number of items produced is found by adding up all five recorded output figures, which is 1725. There are five recorded figures, so our mean is found by dividing 1725 by 5, which is 345 2) Now line B also has five recorded figures which add up to 1725, our mean is found by dividing 1725 by 5, which is 345. So the means are the same, but there is a difference between the two sets of figures which we shouldn't ignore.

ACTIVITY 2
Time guide
1 min

Write down one clear difference between the two sets of figures.

* _____

_____

_____

We can see that line B's figures cover a larger range of values than do line A.

SELF CHECK 10 Note the ranges of:
Time guide
1 min

Line A range = * _____

Line B range = * _____.

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. We defined the range as being the difference between the largest and smallest value in a set of figures. The range for line A is therefore 500 - 150, which is 350. The range for line B is 560 - 90, which 470. There is quite a difference here.

In other words, although the means of the two lines are the same, the range for line B is much larger than that for line A. We need an average and a measure of spread to effectively describe and compare two sets of figures.

Before we move on to Part B of this unit, let's summarise our work so far.

8 Summary

In Part A we've focussed on averages and spread. We looked in particular at

* The Mean - How to calculate it from a set of figures, and from a FREQUENCY DISTRIBUTION TABLE.
* Alternatives to the mean - The MODE and the MEDIAN. We defined these alternatives and invited you to take up an extension to explore them further.
* The Range - How to find the range.
* Alternatives - Ignoring extreme values, or using another measure called the STANDARD DEVIATION.
* The use of averages and ranges - We noted that a good description of a set of figures requires an average and a measure of spread. We used the mean and the range to describe and compare two sets of output figures.

PART B SOME BASIC STATISTICS

1 Introduction

In part A, we considered averages and ranges, and used them to describe and compare sets of figures. In this part, we'll concentrate on some basic statistics which can be used to carry out quick statistical tests. The tests themselves, for the most part, form the material for the rest of the unit. We aim to explore the main ideas behind them, and to point out the dangers. Because of the nature of these quick tests, they may indicate a need for further testing or they may confirm a suspicion. In using tests we must always be aware of the dangers of using statistics at inappropriate times.

REFERENCE 3
Page 60

2 Spotting Rogue Values

When we looked at the mean and the range, we noticed how extreme values can affect our results. If a set of figures contains an extreme value, there are two possibilities. Firstly the value may be genuine, in which case we shouldn't really ignore it. Secondly it may be an error. It is easy enough to record 3.0 as 30, or 1.6 ad 6.1, and so on.

So, if a figure looks out of place, we can apply a test to see if indeed it is out of place.

In some circumstances we may be justified in leaving the figure out of further analysis. But we should always check the source of that out of place figure - it deserves our attention.

Consider the following figures for percentage scrap rates in a particular workplace:

1.6; 3.2; 4.1; 3.2; 29.6; 3.6; 8.6; 7.6

ACTIVITY 3
Time guide
1 min

Do any of these figures look out of place?

* _____

Well it does look as though 29.6 is a lot larger than the other figures which are all below 10.

So we should be curious about this figure of 29.6. Here is a quick test to apply:

Step 1 - Note the range of all the figures. In this case it is 29.6 - 3.2 = 26.4

Step 2 - Note the modified range of the figures, by leaving out the suspect figure of 29.6. So this modified range would be 8.6 - 3.2 = 5.4.

Step 3 - Divide the full range by the modified range.

In this case the full range is 26.4.

The modified range is 5.4.

$\frac{26.4}{5.4}$ is over 4 but under 5, so say 4½.

Step 4 - If step 3 gives a result more than 2, then we can say that the figure which looks out of place should not be included as there is every chance it is out of place.

We were right to be curious about the 29.6 figure. In such cases as this, we should check out the possible reasons for this figure. It could be an error or it could be that the machinery needs attention or the operator needs a little help - perhaps more training.

WE CANNOT AFFORD TO IGNORE ANY OUT-OF-PLACE VALUE

even if our test shows it to be out of place. We should investigate the value and its source.

SELF CHECK 11 Is there an odd figure in the following set of figures?
Time guide
2 mins 1.4; 1.6; 2.9; 3.6; 8.1; 13.4

Fill in the blanks below.

Step 1 - Take the largest figure from the smallest - full range:

* _____

Step 2 - Take the largest figure from the smallest leaving out our suspected odd one out - modified range:

* _____

Step 3 - Divide the full range by the modified range:

* _____

Step 4 - Is the answer greater than two?

Yes or No * _____

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. 13.4 looks out of place as it is the only one over 10. But we don't reject it yet. We'll go through our 4 steps: Step 1: 13.4 - 1.4 is 12, Step 2 8.1 - 1.4 is 6.7, Step 3: 12 divided by 6.7 is just less than 2. Step 4: 12 divided by 6.7 is less than 2 so it seems that the figure 13.4 is genuine and should not be removed from any further calculations. Naturally as it is the highest figure in the set we will still give it some attention and try to see what is causing such a high figure.

Before leaving this section let's recall one of our earlier ideas.

The great value of the mean and the range is that together they can describe or represent data. But we also know that one extreme value in our data can easily distort the range and give a false picture. That's why we developed a quick test to see if it is reasonable to ignore any extreme figure.

By the same token an extreme figure can also distort the mean, though in this case the distortion isn't as great. However, if our quick test shows that an extreme figure should be ignored as far as the range is concerned, then we can also ignore this extreme figure when calculating the mean.

So we can see that even simple statistics like the mean and the range need careful attention - this attention will improve the ability of these statistics to represent the data.

3 The Square Root

As the square root is an often used figure in many statistical calculations and tests we should remind ourselves of what it is before going any further.

2 is the square root of 4 because:

2 x 2 is 4

and

3 is the square root of 9 because:

3 x 3 is 9

So:

THE SQUARE ROOT OF A NUMBER WHEN MULTIPLIED BY ITSELF GIVES THAT NUMBER.

We can easily find a square root using a calculator - the square root button has the symbol $\sqrt{\ }$ marked on it. $\sqrt{\ }$ is a way of saying square root.

Without a calculator it's more difficult. However, for our quick tests we only need to be approximate, so we use commonsense to make a sensible guess.

If we wanted the square root of 10, then we know 3 X 3 = 9 and 4 X 4 = 16 therefore $\sqrt{10}$ is more than 3 and less than 4. So a sensible guess is 3¼.

And $\sqrt{20}$: 4 X 4 = 16; 5 X 5 = 25, so $\sqrt{20}$ is just about between 4 and 5, say 4½.

SELF CHECK 12
Time guide
3 mins

What is the square root of the following numbers?

The square root of 36, or $\sqrt{36}$ is * _____.

The square root of 49, or $\sqrt{49}$ is * _____.

Estimate the square root of the following:

$\sqrt{70}$ = * _____.

$\sqrt{30}$ = * _____.

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. 6 when multiplied by itself gives 36. So, 6 is the square root of 36. 7 when multiplied by 7 gives 49 so the square root of 49 is 7. Now to the estimates, 8 x 8 = 64 and 9 x 9 = 81 so $\sqrt{70}$ is roughly $8\frac{1}{2}$, 5 x 5 = 25 and 6 x 6 = 36 so $\sqrt{30}$ is roughly $5\frac{1}{2}$.

Square roots are important in our tests. The standard deviation is a basis for so many of the quick tests in this unit that we can't afford to ignore it. So, let's take a quick look at the standard deviation.

4 The Standard Deviation

From our earlier work we know that the standard deviation is an alternative and very useful measure of spread.

We don't need to know how to calculate the standard deviation, but an awareness of what it is will help us understand the analytical tool we'll be using later on in the unit. However if you do want to learn how to calculate this statistic then take up this extension. Unlike our other extensions this one is complete in itself as you'll see when you look it up.

EXTENSION 2
Page 56

We have said that the larger the value we got for a measure of spread, the more spread out our values are. The standard deviation is a measure of spread, so the larger it is, the more spread out our data will be.

Let's try to settle what this means without getting too deeply into difficult and, for us, unnecessary areas of complicated statistics.

Firstly let's see what we mean by a Normal Distribution. One simple way of thinking about this is to say that a Normal Distribution occurs where a set of values tend to cluster around the mean in a particular way.

For example, the mean height of adults in one country could be 5'9". It's obvious that there will be some people shorter than this and some taller - however we really know a bit more than this:

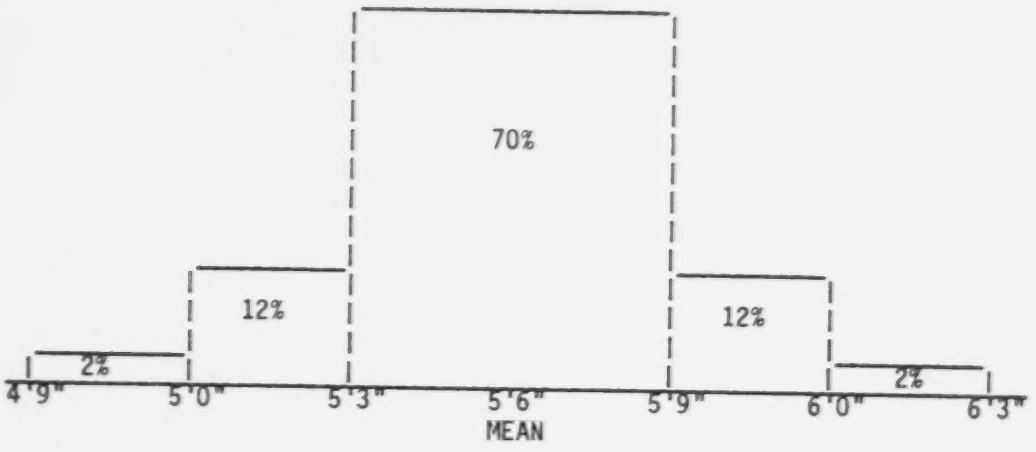
A lot of adults will BE CLOSE to 5'6"

Quite a number will be a BIT FURTHER away from 5'6"

A few adults will be a LOT FURTHER away from 5'6"

We could also put some figures to our guesses, for example we could say that 70% of people will be within 3" of 5'6".

Although they are only guesses we can put this data into a diagram:



So this height data is clustered around the mean of 5'6". Our main point is that 94% (70% + 12% + 12%) of these adults are between 5'0" and 6'0" tall.

What about our standard deviation then? Well we could say that 3" is our standard deviation. We can easily see how this can give us a measure of spread.

Now all this is designed only to give a rough idea of the relationship between a Normal Distribution and its Standard Deviation. We don't need to know more than this. The main point about this relationship is that:

95% of the data in a Normal Distribution will be between:

a point 2 standard deviations below the mean

and

a point 2 standard deviations above the mean

You'll recall we used the number 2 earlier on to spot an unusual figure - perhaps a rogue. Well this is where our figure 2 comes from - from the fact that in a normal distribution 95% of all the data lies within 2 standard deviations of the mean.

We therefore gamble that, if a figure which we are checking on is more than 2 standard deviations from the mean, then it is a 'rogue' and needs checking out. It's a pretty good gamble - the odds are on our side - we will be proved right 95% of the time or 19 times out of 20. So we'll be right most of the time.

So our standard deviation has at least two uses to us here.

IT HELPS US TO LOCATE A POSSIBLE "ROGUE" FIGURE IN DATA AND IT GIVES A MEASURE OF SPREAD.

So, let's compare some sets of figures to see which are more widely spread.

ACTIVITY 4
Time guide
2 mins

The list below gives the standard deviations for 8 sets of figures.

0.8; 1.4; 0.6; 1.9; 0.4; 2.0; 0.8; 1.6

a) Which figure indicates the widest spread * _____.

b) Which figure indicates the narrowest spread * _____.

The distribution which has the widest spread is the one with the standard deviation of 2.0. The distribution which has the narrowest spread is the one with the standard deviation of 0.4.

Let's leave it there and move on to see how to use what we know in practice.

5 An Acceptable Level For Testing

Most of what follows in this unit is based upon a unique relationship between the standard deviation and a Normal Distribution.

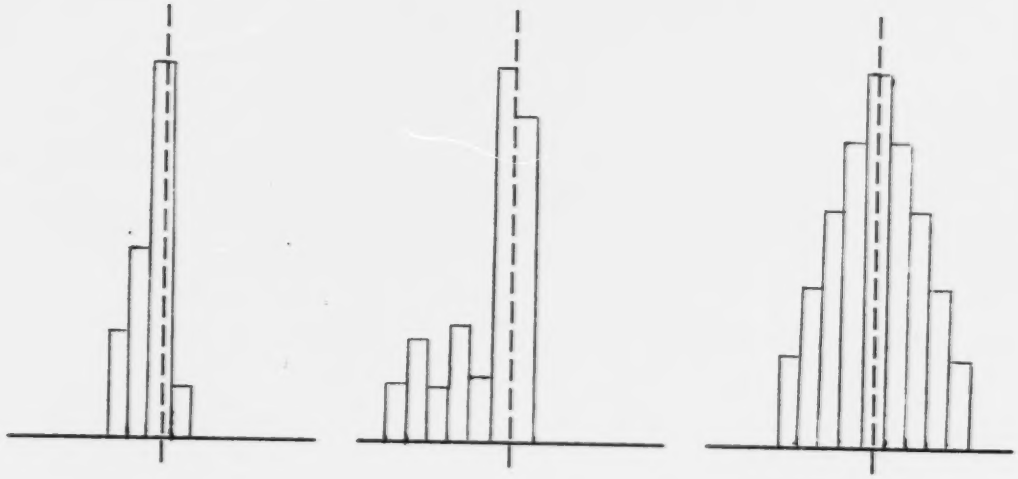
If we could plot every possible value on a graph for, say:

- numbers of operators and output
- number of machines and defects
- number of people and shoe sizes
- number of sales of washing machines by washing machine salesmen

and then join them up, the result could be a normal distribution.

You'll recall we said earlier that in a Normal Distribution the values tend to cluster around the mean. We can add, too, that the pattern of values above the mean looks very much like the pattern of values below the mean.

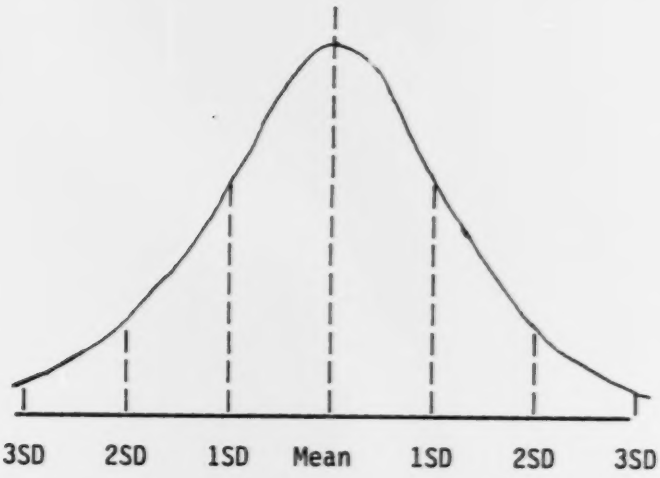
SELF CHECK 13 Put a tick beside the diagram which shows a normal distribution:
Time guide
2 mins



The answer to this self check is here in the following words, so you'll need to check out your answers carefully. The most correct response is C. As we can see the data is clustered around the mean and the data below the mean is a mirror image of the data above the mean. Diagrams A and B are not like this.

We should also point out that we don't always have data that gives such a perfect shape as in diagram C - as long as it is approximately this shape we're OK to assume it is a Normal Distribution.

Let's look at an example - you'll notice that this normal distribution is drawn as a curved "bell-like" diagram - this is the usual way of drawing it.



SD = Standard Deviation

SELF CHECK 14
Time guide
1 min

Recalling what we said earlier about the relationship between the normal distribution and the standard deviation,

what percentage of the values lies between a point 2 standard deviations above the mean and a point 2 standard deviations below the mean?

* _____

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. 95% - a figure you'll recognise as we used it earlier.

SELF CHECK 15 Complete each sentence with a suitable word or words.
Time guide
2 mins

1 Much of our statistical work is based on the knowledge that many groups of values form a _____ distribution.

2 The range is a measure of the spread of data, so is the _____.

3 There is a special relationship between the _____ and the _____.

4 95% of all values in a normal distribution lies between _____ standard deviation above the mean and _____ standard deviation below the mean.

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. 1) normal, you'll remember the bell-shaped diagram. 2) standard deviation, in many ways this is a more useful statistic. 3) normal distribution and standard deviation, with these two statistics we can generate a lot of information about the values. 4) two and two, by operating at two standard deviations we can take some calculated gambles.

I think that we have now established the basis for a useful statistical test that has many practical applications.

The two basic parts of our test are:

the number 2 - 2 standard deviations above and below the mean includes 95% of our values

- it gives us gambling odds of 95 out of 100 or 19 out of 20 that we'll be right.

$\sqrt{\ }$ - the square root sign

We won't go into our test any further at this stage - we'll do this in Part C.

6 Summary

* To spot rogue values in a range of values we used a 4 step method to check on the suspect (rogue) value.

  Step 1 - calculate the full range of all the values

  Step 2 - calculate the modified range by excluding the suspect rogue

  Step 3 - divide the full range by the modified range

  Step 4 - if our result from step 3 is more than 2, then our suspect is probably a rogue and should be checked out.

* Square root - the symbol $\sqrt{\ }$ means square root

  so $\sqrt{16} = 4$ and $\sqrt{25} = 5$

* Standard deviation is a special measure of spread more useful than the range

  - when applied to the normal distribution we know that 95% of all values lie between 2 standard deviations above the mean and 2 standard deviations below the mean. So we can also use the standard deviation to spot a "rogue" value.

* Normal distribution - we decided not to go too deeply into this but we did note that in a normal distribution values tended to cluster around the mean.

  We did note that much of our data at work would form a normal distribution.

* Acceptable level for testing or our gambling odds

  by adopting 2 standard deviations as our limit we give ourselves a 95% chance of being right - and of course a 5% chance of being wrong!

* Our useful statistical test uses:

  this figure of 2

  and the $\sqrt{\ }$

  we decided to leave it there and develop this test in Part C. Let's do this now.

PART C OUR BASIC STATISTICAL TEST

1 Introduction

Here we will take up the ideas we've developed earlier and establish our statistical test. This test, we will find, gives us a sound base for making decisions in that we'll be right 19 times out of 20. We will apply this same test to a number of different situations, so proving to ourselves just how useful our test is. But first a note of caution. We can use our analytical tool to guide us towards better decision making, but as Capey and Carr stress no technique can make decisions for us.

REFERENCE 2
Page 60

2 Our Statistical Test

Let's establish the formula for our test through an example. There are occasions, as we have seen, when a figure looks out of place and can be identified as a "rogue" - as a figure which shouldn't be included. However on other occasions we may find that our suspect "rogue" is not a rogue after all and should be included as part of the data.

ACTIVITY 5
Time guide
5 mins

Here's a short incident:

The percentage of defective components produced by the operators in a section making components averages 8%. One of the operators appears to be producing a significantly different percentage of defects than the others. On twenty occasions his output was checked, and his percentages of defects were

7.5, 8.1, 8.5, 9.0, 8.0, 8.2, 8.4, 8.0, 9.0, 9.1
7.9, 8.0, 8.0, 8.4 8.6, 9.0, 7.5, 7.6, 8.2, 8.0

Perhaps we shouldn't really say it, but have we got a "rogue" operator here! We are suspicious but we need to confirm or otherwise our suspicions.

* So is this operator a "rogue"? YES/NO

* Give a short reason for your answer:

_____

_____

_____

We could argue that he produces a significantly higher percentage of defects as his percentages are not very often below the 8% average.

On 11 occasions he produced more than 8% of defects.

On 5 occasions he produced 8% of defects.

On only 4 occasions did he produce less than 8% of defects.

Another approach might have been to calculate his mean percentage of defects. If this method was used, the mean percentage of defects would have been 8.25%, or the total of the values divided by the number of values. In other words, 165 divided by 20 which is 8.25%.

As 8.25% is higher than our mean of 8% it might be concluded that the operator does produce a higher percentage of defects.

The problem here is that 8.25% is not very much higher than 8.0%.

The line of thought we have just described seems reasonable but we should be certain in our own minds about our judgement before taking action. We can test his results now that we know that

- 95% of all values for defects should fall within two standard deviations above the mean and two standard deviations below the mean.

We don't need to calculate the mean of our sample nor its standard deviation. Fortunately, there is an easier way than this - we compare the number of values over 8% to the number below 8%.

Here is the test.

- count the number of values higher than our mean of 8.0%
- count the number of values lower than our mean of 8.0%
- find the difference between the two.

Note: we ignore the values equal to the mean of 8.0%.

ACTIVITY 6
Time guide
2 mins

Fill in the blanks in the spaces provided to complete the first part of our test on the operator's results.

His percentages of defects on the twenty occasions were:

7.5, 8.1, 8.5, 9.0, 8.0, 8.2, 8.4, 8.0, 9.0, 9.1
7.9, 8.0, 8.0, 8.4, 8.6, 9.0, 7.5, 7.6, 8.2, 8.0

The number of values higher than our mean of 8.0 is *____.

The number of values lower than our mean of 8.0 is *____.

The difference between them is *_____.

The number of values higher than our mean of 8.0 is 11.

The number of values lower than our mean of 8.0 is 4.

The difference between those results of 11 and 4 is 7.

We saw in the last section that 95% of all values could be expected to fall within two standard deviations of the mean. Because, for our test, the standard deviation is the square root of the sum of our differences, we can estimate our standard deviation from our "differences" - in our case the differences are 11 and 4.

So we can now work out the "Sum" of our "Differences".

ACTIVITY 7
Time guide
2 mins

Fill in the blanks below:

The sum of our differences is * _____.

The square root of this number is * _____.

The sum of our "differences" is 11 + 4. Which is 15. The square root of 15 is just under 4.

Now we have identified and defined the "square root of the sum of those differences" we can note a shorthand way of writing this :-

$\sqrt{\text{sum}}$; in our example is $\sqrt{11 + 4}$ = nearly 4

We can now introduce our 2 standard deviations figure, so our formula builds up to:

$2\sqrt{\text{sum}}$; in our example 2 x nearly 4 = nearly 8

So what do we do with this result? How do we use it?

Well, by working out $2\sqrt{\text{sum}}$ we are finding a value which will not be exceeded by 95% of all normal values. This reminds us of our work on the normal distribution.

So if our actual difference - the one we are concerned about - is less than this $2\sqrt{sum}$, then it comes within this 95% of values and is quite normal - in other words it is not a "rogue". It sounds a bit complicated but it comes easier with practice!

Let's summarise out test:

If the Actual Difference is less than $2\sqrt{sum}$ then the result IS NOT a "rogue".

If the Actual Difference is more than $2\sqrt{sum}$ then the result IS a "rogue".

SELF CHECK 16
Time guide 3 mins

Let's return to our example and put all our calculations together. Fill in the blanks:

Number of values higher than our mean of 8 is * _____.

Number of values lower than our mean of 8 is * _____.

The actual difference between these values is * _____.

The sum of these differences is * _____.

$2\sqrt{sum}$ is * _____.

The actual difference is * _____ than * _____ therefore the set of values _____ a "rogue" set.

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. Number of values higher than our mean of 8 is 11. Number of values lower than our mean of 8 is 4. The actual difference between these values is 7. The sum of these differences is 15. $2\sqrt{sum}$ is nearly 8. The actual difference is less than nearly 8 therefore the set of values is not a "rogue" set. So we have no real grounds for action though we would be wise to continue to check up on his future performance.

SELF CHECK 17 Here are two more examples for practice:

Time guide
10 mins

1 Here's a set of defect percentages produced by another operator:

8.4, 8.0, 8.1, 7.9, 8.3, 8.1, 8.2, 8.0, 8.4, 8.1
8.2, 8.3, 8.1, 7.6, 8.4, 8.5, 8.6, 8.1, 8.2, 8.3

If our mean percentage of defects is still 8.0% do these new results suggest that our operator is producing too high a percentage of defects.

2 Let's assume that the supervisor was able to use our test to show that the percentage of defects of a particular operator was far too high. The operator then blames the machine. So to run a check another operator was put on this machine and the following results were obtained.

8.0, 7.9, 8.1, 8.0, 7.9, 8.2, 8.0, 8.1 8.1
8.4, 7.8, 8.0, 7.9, 8.2, 8.4, 8.0, 7.8, 8.2

Given that the overall mean percentage of defects is 8.0, is the machine at fault?

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. 1) The number of values higher than our mean of 8.0 is 16. The number of values lower than the mean of 8.0 is 2. The actual difference is 16 - 2 which is 14. The sum is 16 + 2 which is 18. The square root of the sum is $\sqrt{18}$ which is just over 4. Twice the square root is over 8. Our actual difference of 14 is well above 8. This tells us with 95% certainty that the operator does produce too high a percentage of defects. 2) The number of values higher than our mean of 8.0 is 8. The number of values lower than our mean of 8.0 is 5. The actual difference between them is 8 - 5 which is 3. The sum is 8 + 5 which is 13. The square root of the sum is $\sqrt{13}$ which is about 3½. $2\sqrt{13}$ or 2 x 3½ is 7. The actual difference of 3 is well below 7, which tells us with 95% certainty that the machine does not produce an unusually high rate of defects. So the machine is not at fault - the operator is.

This test looks like being very useful and, with a little practice, is easy to do. We can also extend the use of this basic test to other data.

3 Comparing Two Sets of Figures

It can often be useful to compare two sets of figures to see if they differ significantly from each other. By significantly we mean outside the 95% range of 2 standard deviations above and 2 standard deviations below the mean.

ACTIVITY 8
Time guide
5 mins

Note two examples of sets of figures we might want to compare. I have entered one as an example:

* Scrap rates in each of 2 similar factories, sections or work teams.
* _____
* _____

Typical responses might have included:

* Output figures in each of two similar factories, sections or workteams.
* Earnings levels in each of two similar businesses.
* Weekly takings in each of two similar retail organisations.

We could list many more kind of figures which we could compare.

Let's look at two competing and similar supermarkets. The table below shows the weekly takings of each. We want to know if the takings of one differs significantly from the takings of the other.

Weekly takings for each of two supermarkets

Week	Supermarket A £	Supermarket B £
1	4600	4900
2	4700	4900
3	4900	4850
4	5000	4950
5	4900	5000
6	5000	5100
7	4800	4900
8	4950	5000
9	4900	5100
10	4950	5010
11	4900	5000
12	4960	4990

We can use our statistical test to run a check on these two supermarkets.

Our first step is to compare the takings of the two supermarkets week by week. We do this by writing in the answer to two questions:

How many times did supermarket A take more money than supermarket B?

How many times did supermarket B take more money than supermarket A?

We will then have our two numbers to compare in exactly the same way as before. By applying our test we can see whether one supermarket takes significantly more money than the other.

We know enough about our test to do this.

SELF CHECK 18
Time guide
3 mins

Look back to the table of supermarket takings to get the two numbers we need and then complete our statistical test.

Fill in the blanks below:

Number of times when A is greater than B * _____

Number of times when A is less than B * _____

Difference between these two numbers * _____

Sum of these two numbers * _____

$2\sqrt{\text{sum}}$ * _____ approximately

Is difference greater than $2\sqrt{\text{sum}}$? * _____

Is the difference between the takings significant? * _____

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. The number of times when A is greater than B is 2. The number of times when A is less than B is 10; The difference between these two numbers is 8; The sum of these two numbers is 12; $2\sqrt{\text{sum}}$ is 7; Is difference greater than $2\sqrt{\text{sum}}$, yes; Is the difference between the takings significant, yes.

Because Supermarket B took more money than A on most weeks, then we can say that Supermarket B took significantly more money than A.

So we have extended the use of our test to cover another situation. This is good, we use the same method and the same calculations, so we won't overstrain our memory.

SELF CHECK 19 Let's take a look at another example
Time guide
5 mins

THE OUTPUT OF 2 WORK TEAMS OVER A PERIOD OF 10 WEEKS

WEEK	WORKTEAM A	WORKTEAM B
1	1600	1660
2	1700	1660
3	1800	1790
4	1750	1790
5	1810	1785
6	1810	1750
7	1810	1765
8	1900	1800
9	1950	1945
10	1975	1950

Fill in the blanks:

Number of weeks when A was greater than B * _____

Number of weeks when A was less than B * _____

Difference between these two numbers * _____

Sum of these two numbers * _____

$2\sqrt{\text{sum}}$ * _____ approximately

Is difference greater than $2\sqrt{\text{sum}}$? * _____

Note your conclusion based on this test:

* _____

_____

_____

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. Number of weeks when A was greater than B is 8; Number of weeks when A was less than B is 2; Difference between these two numbers is 6; Sum of these two numbers is 10; $2\sqrt{\text{sum}}$ is 6 approximately; Is difference greater than $2\sqrt{\text{sum}}$, no. So we would conclude that there is no real or significant difference between the performance of these two workteams.

4 Sample Size

By sample size we mean the number of values we are working with. In the case of the operator who we suspected was producing too high a percentage of defects, the number of figures is the sample size. But where we were comparing one supermarket with another, the sample size is the number of weeks we were looking at.

The question we need to ask is: "is our sample size sufficient for the conclusion we made?"

We'd clearly agree that checking the takings of one supermarket to another on only one day would tell us little about the weekly or monthly comparison. So how many days must we check?

Let's not labour the point - we'll keep it simple by saying that we need at least 10 figures to decide. There is a statistical way of deciding on the proper sample size - if you want to check this out then you can take up this extension.

EXTENSION 3
Page 59

5 Summary

* Our basic statistical test compares:

  $2\sqrt{\text{sum}}$ with the actual difference

* If actual difference is greater than $2\sqrt{\text{sum}}$, then the difference is real and significant and should be investigated.

* Sample size should be at least 10

  - there is a method for setting the sample size in extension 3.

* Our test gives us 95% certainty in our decisions.

PART D RELATIONSHIPS AND TRENDS

1 Introduction

Having seen how our basic test $2\sqrt{\text{sum}}$ can be applied to a variety of situations, let's extend its use ever further. Here we will apply the same test to data which at first may look a little more complicated. We will check to see if two sets of data are related in some way, for instance we would expect eye accidents to fall as more and more people used their safety goggles. We will also see how we can check out whether our data on productivity is showing a consistent rise,in other words showing a rising trend.

2 Relationships in Data

Much of the data in our work places is related. Another way of putting this is to say that much of our effort at work is designed to improve things, like spending money on safety or advertising. In these cases we would expect to see a fall in accidents and an increase in sales. Let's look at some data which may be related to other data.

ACTIVITY 9
Time guide
1 min

List two instances of data which may be related:

eg. Hours worked in a shift and the spoilage rate

* _____

* _____

Typical responses might include:

* Hours of training and productivity
* Use of safety equipment and accident rate
* Use of mechanical aids and back complaints
* Output and defective rates

Many other pairs of items could be listed where, like the ones here, a change in the first item would be matched with a change in the second. Let's see how we can use our test to check on data like this.

Let's assume that 12 operators work a varying amount of their hours on a particular operation. We check out their average defect rate over these periods. We do this because we suspect that working longer hours seems to be creating more defects.

Here are the hours and defects for each operator:

Operator	Hours	Defects	Operator	Hours	Defects
a	7.5	2.1	g	11.1	2.9
b	8.5	3.6	h	11.3	2.5
c	12.2	3.9	i	9.5	1.9
d	10.1	4.1	j	9.7	3.6
e	12.4	3.6	k	8.0	2.9
f	12.5	3.0	l	8.0	2.0

ACTIVITY 10
Time guide
3 mins

List this data in an ascending order of hours, making sure to keep the data in its original pairs. Then calculate the means for the hours and for the defects.

Operator	Hours	Defects
a	7.5	2.1
l	8.0	2.0
k	8.0	*___
*___	*___	*___
*___	*___	*___
*___	*___	*___
*___	*___	*___
*___	*___	*___
*___	*___	*___
*___	*___	*___
*___	*___	*___
*___	*___	*___
Means	*_____	*_____

I don't need to reproduce the list here for the moment, but the means are: for the hours, 120.8 divided by 12 and is approximately 10.1; for the defects, 36.1 divided by 12 and is approximately 3.0.

We'll develop our next step through another activity.

ACTIVITY 11
Time guide
2 mins

Check back to our listed data and our two mean values to answer the following:

The number of pairs where both values are ABOVE their mean is * _____.

The number of pairs where both values are BELOW their mean is * _____.

The number of pairs where one value is ABOVE its mean and the other is below its mean is * _____.

Note - we have therefore ignored any pair where one or other or both, are equal to its mean.

We should be able to agree that:

The pairs where both values are ABOVE their mean are c; e: and the pairs where both values are BELOW their mean are a; i; k; l: this gives a total of 6 pairs of values.

The pairs where one value is ABOVE its mean and the other BELOW its mean are b; g; h; j: and is 4 values.

Note - we have ignored values d; f; and is 2 values.

We compare our 6 pairs of values to our 4 pairs of values by using our $2\sqrt{sum}$ test.

SELF CHECK 20
Time guide
2 mins

Recall our test by completing this calculation for the example we have just looked at:

$2\sqrt{sum}$ = $2\sqrt{__}$ = 2 x __ = * _____

difference = __ - __ = * _____

$2\sqrt{sum}$ is * _____ than our difference of * _____

so there is * __ _ real relationship between the two sets of data.

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. $2\sqrt{sum}$ is $2\sqrt{6 + 4}$ or $2\sqrt{10}$ which is just over 2 x 3 or 6; the difference is 6 - 4 or 2; $2\sqrt{sum}$ is more than our difference of 2 so there is no real difference in our data.

You'll recall why we ran this check, it was because we suspected that the longer hours were creating higher defects. We now know that this is not the case and we were wrong to suspect it. It's doing another check next week, it doesn't take long to do.

Here's another set of figures - they give the average spoilage rates against the average training hours given by the supervisor in a busy office. So we are really checking up on whether the training is helping spoilage rates to fall.

Training Hours	6.2	6.7	6.2	6.8	6.3	6.9	6.3	6.1	7.0	6.5	6.8	6.4
Spoilage Hours	4.3	2.6	3.8	2.9	3.3	2.6	3.6	3.5	2.4	3.3	3.4	3.9

SELF CHECK 21
Time guide
3 mins

Complete the test to answer the question - "does training have an effect on the spoilage rate"?

Fill in the blanks:

Mean spoilage rate is *____

Mean training hours is *____

Number of pairs where both are ABOVE their mean and number of pairs where both are BELOW their mean is *____ + *____ = *____

Number of pairs where one value is ABOVE its mean and the other BELOW its mean is *____

$2\sqrt{sum}$ = $2\sqrt{__}$ = *____

Difference = *____

$2\sqrt{sum}$ is *____ than the difference of *____, so training hours *____ have an effect on spoilage rates.

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. Mean training hours is 6.5 (just a little bit over to be exact); mean spoilage rate is 3.3; number of pairs where both are above the mean plus the number of pairs where both are below the mean is 1 (1 + 0); number of pairs where one value is above its mean and the other value is below its mean is 9; $2\sqrt{sum}$ is $2\sqrt{10}$ or just over 6; Difference is 8 (9 - 1); $2\sqrt{sum}$ is less than the difference of 8, so training hours does have an effect on spoilage rates.

But, just a minute! We want to know more than this don't we?

ACTIVITY 12
Time guide
2 mins

What else do we want to know? Jot down your ideas:

* _____

_____

Well, we really want to know if the spoilage rates will fall when we increase our training hours. We haven't really checked on this yet we hope this is so, but we shouldn't assume this. Let's check it out properly.

Here's our original list but this time, we have listed the training hours in an ascending order; we have also kept our original pairs intact:

Training hours	Spoilage rates
6.1	3.5
6.2	3.8
6.2	4.3
6.3	3.3
6.3	3.6
6.4	3.9
6.5	3.3
6.7	2.6
6.8	2.9
6.8	3.4
6.9	2.6
7.0	2.4

The line drawn across our data represents our mean of training hours which you'll recall was just over 6.5.

I think we'd accept that higher training hours should reduce spoilage. So we check to see if the spoilage rates with higher than the mean training hours are less than those achieved with lower training hours.

We compare the mean spoilage rate above our line with the mean spoilage rate below our line.

SELF CHECK 22 Calculate these two means to check whether training reduces spoilage rates:

Time guide 2 mins

with less than the mean training hours,

the mean spoilage rate is *____.

With more than the mean training hours,

the mean spoilage rate is *____.

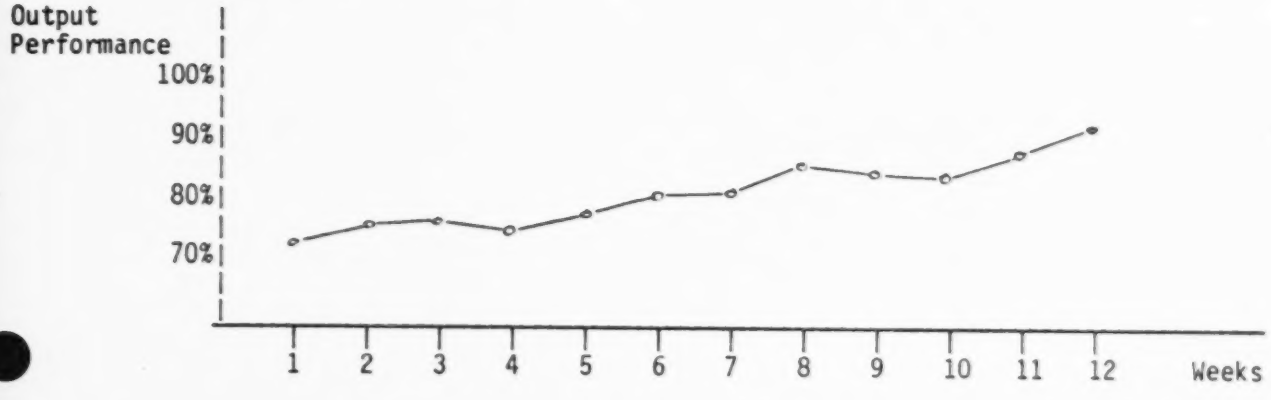
The answer to this self check is here in the following words, so you'll need to check out your answers carefully. With less than the mean training hours the mean spoilage rate is approximately 3.7 and with more than the mean training hours is approximately 2.8.

We can now say with 95% certainty that increasing training hours does reduce spoilage rates. This is valuable information extracted from the data, and once again we've done this with our basic test, so we don't have a lot to remember. But we've not finished yet, our test has yet another use.

3 Spotting a Trend

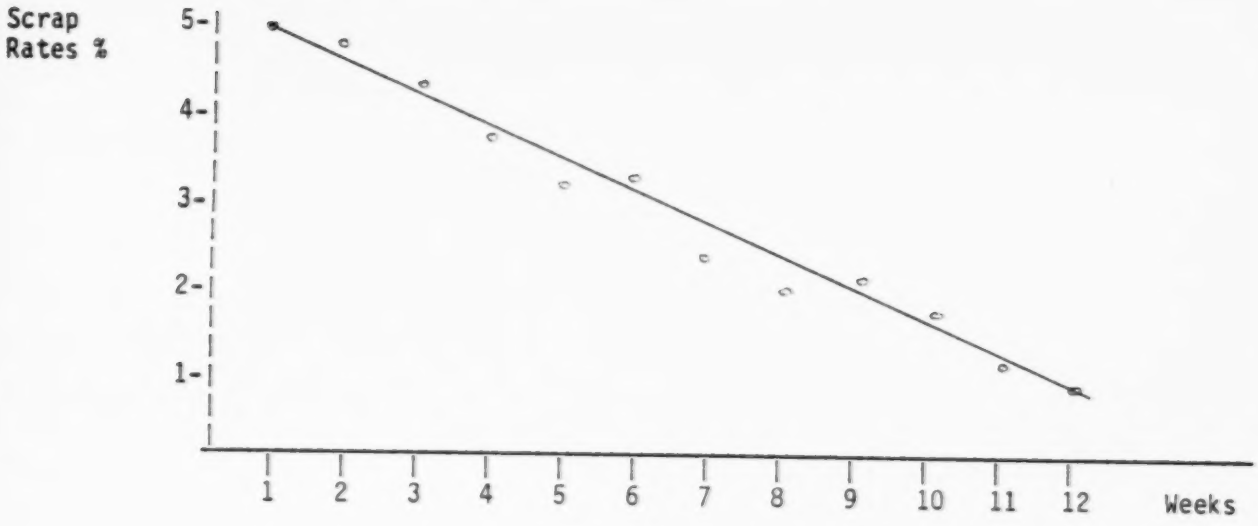
Now a trend is described as a general upward or downward movement of data. A diagram will help to show what this means and it will also show the idea that our test is based on.

Here is a trend of improving output rates following the introduction of a new method:



41

Here is a trend of reducing scrap rates but this time we've drawn a straight line through the data to show the underlying downward trend:



OUR TEST CAN BE APPLIED TO THE FACT THAT THE DATA IN THE LATER WEEKS IS CONSISTENTLY DIFFERENT TO THE DATA IN THE EARLIER WEEKS.

For example consider the scrap rate trend; the values for weeks 9, 10, 11, 12 are all less than the values for weeks 1, 2, 3, 4. We could therefore apply our test to these values. Let's see how we can check for a trend but without having to draw a graph:

Here are sales figures in tons for a certain company over a six year period.

Quarter	Yr 1	Yr 2	Yr 3	Yr 4	Yr 5	Yr 6
1	71	72	73	76	84	89
2	89	90	90	91	94	96
3	94	97	101	103	110	108
4	99	102	109	108	98	100

The question we could ask ourselves about these figures might be "Are sales rising?"

Another way of putting this question is to ask ourselves whether the last 12 figures are higher than the first 12 figures.

Let's see how we can apply our basic test by first of all averaging our figures to get the data to test.

We list our first 12 figures in a column then list our second 12 figures in a second column, being careful to list all our figures in the proper time sequence or chronological order:

First Twelve	Second Twelve
71	76
89	91
94	103
99	108
72	84
90	94
97	110
102	98
73	89
90	96
101	108
109	100

ACTIVITY 13
Time guide
2 mins

Look at the PAIRS of values in our list and answer the following 2 questions:

a How many times is the second figure HIGHER than the first *____

b How many times is the second figure LOWER than the first *____

Ignore any pair where the figures are the same.

By carefully checking we have:

Second figure is higher 10 times and second figure is lower 2 times.

From a careful look at our data we can probably see that the value 10 suggests a RISING trend and the value 2 suggests a FALLING trend.

This is the data for our familiar $2\sqrt{\text{sum}}$ test.

SELF CHECK 23 Complete the test by filling in the blanks:
Time guide
3 mins

The number of values suggesting a rising trend is *_____

The number of values suggesting a falling trend is *_____

The difference between these numbers is *_____

The sum of these numbers is *_____

$2\sqrt{\text{sum}}$ is *_____ approx.

$2\sqrt{\text{sum}}$ is *_____ and is *_____ than the difference of *_____, so there *_____ a trend.

The answer to this self check is here in the following words and calculations, so you'll need to check out your answers carefully. Here's the completed test. The number of values suggesting a rising trend is 10. The number of values suggesting a falling trend is 2. The difference between these numbers is 8. The sum of these numbers is 12. $2\sqrt{\text{sum}}$ is 7 approximately. $2\sqrt{\text{sum}}$ is 7 and is less than the difference of 8, so there is a trend.

Because 10 of the values suggested a rising trend and only 2 suggested a falling trend, we can be 95% sure that there is a rising trend in the sales across these 6 years.

Here's another example to finish off this part of our unit.

SELF CHECK 24
Time guide
5 mins

You'll be quite familiar with our method by now so here are the figures for your list. Complete the test in the space below and answer the two questions:

Quarter	Yr 1	Yr 2	Yr 3	Yr 4	Yr 5
1	64	61	62	63	64
2	61	62	64	65	64
3	61	63	65	66	
4	59	60	60	62	

First Nine	Second Nine
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

Is there a TREND? YES/NO

If so, is it RISING/FALLING?

Here is a complete check.

First Nine	Second Nine
64	64
61	65
61	60
59	63
61	65
62	66
63	62
60	64
62	64

The number of times where the second figure is higher is 6; The number of times where the second figure is lower is 2; The difference between these numbers is 4; The sum of these numbers is 8; $2\sqrt{8}$ is nearly 6; $2\sqrt{8}$ is 6 and is greater than the difference of 4; so there is no real trend.

We have now seen just how useful our $2\sqrt{\text{sum}}$ test is - it has many applications.

After some practice you'll find you are able, in some cases, to do this test "in your head" without having to write it all down. When you can do this you really will have a powerful aid to decision taking - you'll know what to decide and be confident in your decision; and you will be right many, many more times than you will be wrong.

4 Summary

* This part of our unit extended the use of our test to see if:

- two sets of data were related

- a trend existed in chronological data.

EXTENSION 4
Page 59

There are other quick tests which can be easily carried out using graphical methods. We've not looked at these in this unit, but perhaps you'd like to take up our last extension and explore them yourself.

PART E PERFORMANCE CHECKS

Time guide 60 mins

1 End Check

You may review your understanding of the unit as a whole completing this group of questions. **DON'T GUESS THE ANSWERS** - if you're not sure, flick back into the unit and check before responding.

If you are still unsure then put a question mark against your response before proceeding.

TRUE/FALSE

Respond to each statement by writing TRUE or FALSE in the space provided.

1 The range is a measure of spread. _____

2 The mean can represent a set of data. _____

3 The range is not affected by extreme values. _____

4 The standard deviation is a measure of spread. _____

5 By removing an extreme value the range can become more realistic. _____

6 The mode and median, like the mean, can represent a set of data. _____

7 The mean is the sum of all values divided by the the number of values. _____

8 The range is a more useful statistic than the standard deviation. _____

COMPLETION

Respond by completing each sentence with a suitable word or words.

9 We used the Greek symbol 'sigma', $\sum$, to mean _____.

10 The table we used to help our calculations of mean was a _____ table.

11 The range _____ affected by extreme values.

12 The _____ is an alternative to the range and is more useful.

13 A set of data can be well described by giving the mean and _____.

14 The _____ of 16 is 4.

15 We can check on a suspect rogue figure by comparing the full range to the range which _____ the suspect rogue.

16 Our basic statistical test is based on 2 $\sqrt{*}$_____.

MULTIPLE CHOICE

Respond to each question by putting a mark against the ONE answer which is most correct.

17 The range of this data, 11; 16; 18; 3; 22; 14, is

a 11

b 22

c 19

d 3

18 The mean of this data, 11; 16; 18; 3; 22; 14, is

a 17

b 7

c 16

d 14

19 The square root of 15 is

a nearly 5

b nearly 3

c nearly 4

d just over 4

20 $2\sqrt{24}$ is

a 12

b Just under 10

c Just over 10

d Nearly 4.9

21 The range of two standard deviations above the mean to two standard deviations below the mean will contain:

a 47½% of all values

b 100% of all values

c 95% of all values

d 5% of all values

22 We can use our basic test $2\sqrt{sum}$ to check on:

a the standard deviation

b the significance of a difference

c the range

d the mean

23 If $2\sqrt{sum}$ is 6 and the actual difference is 8 then:

a the difference is not real and significant

b the difference is real and significant

c we are 95% sure that the difference is real and significant

d we can draw no conclusion from the difference

24 When we wish to test whether one set of figures is related to another set of figures, then we begin by:

a calculating the mean of all our figures

b calculating the mean of each set of figures

c seeing how many times one of the figures in the pair is higher than the other

d checking how many times both figures in the pairs are the same

25 We check on the past 20 days output from two machines. We find that the difference in the mean outputs is more than $2\sqrt{sum}$ of these means. We can therefore say:

a we are 95% sure that the difference is real and significant

b the difference is not real and significant

c we are 100% certain that the difference is real and significant

d we are 50% certain that the difference is real and significant

26 We check on the output figures for the past 38 week, by comparing weeks 1 to 19 to those of weeks 20 to 38. We find 13 figures in the later weeks higher than the corresponding ones in the earlier weeks; 3 figures lower; 3 figures the same. We are therefore:

a sure there is a trend

b sure there is not a trend

c sure there is an upward trend

d 95% sure there is an upward trend

END CHECK - Checkout

Audio Tape Side 2 Time guide 35 mins

The answers and explanations for these End Checks are on Side 2 of our Audio Tape. Have your responses with you as you play this tape. As you listen you can check out how well you have done and make notes as you proceed.

2 Tutor Check

Time guide 60 mins

New production methods were introduced in a number of sections with a view to cutting down scrap rates and increasing productivity. Each of these sections were then to be compared with similar sections which had not gone over to the new methods.

The following results were recorded over a 20 week period for a group using the new methods and a group not using them:

Old Methods		New Methods	
Weekly output in tons	percentage scrap	Weekly output in tons	percentage scrap
55	6.2	58	6.1
56	6.2	62	5.9
57	6.1	60	5.9
58	6.2	59	6.0
57	6.2	60	5.6
56	6.1	61	5.4
55	6.0	62	5.6
54	6.0	62	5.5
56	5.9	61	5.3
57	6.1	62	5.2
58	5.8	63	5.4
59	5.8	64	5.6
60	6.1	64	5.5
58	6.1	65	5.4
57	5.9	66	6.0
56	5.8	61	5.6
57	5.9	70	5.5
57	6.0	71	6.1
56	6.0	69	5.4
55	6.1	71	5.3

Comment on these results after applying statistical tests developed in this unit.

Use your test to answer the following questions:

1 Is the difference in mean outputs real and significant?

2 Is the difference in mean % scrap real and significant?

3 Check on the new method to see whether output is rising by checking for a trend.

4 Do you recommend a switch to the new methods? Give the reason for your recommendation.

Don't forget, show all your calculations as well as giving your one sentence answers.

Send to tutor

Now get your answers to your study tutor either by post or handing it over personally. You will receive your tutor's assessment and comment back through the post as soon as can be managed.

3 Work-based Project

Time guide 60 mins

1 Against each of these three tests, make a note of two sets of data from your workplace that the test could be applied to:

a Testing for a difference between two means.

b Testing for a trend.

c Testing for a relationship.

2 Gather some data for one of your two choices against each of these tests and conduct our basic test. State clearly the conclusion you draw from each of the three results - a single sentence for each conclusion is required.

3 Note any action that may prove helpful due to any of these conclusions you have drawn. A single sentence for each action is all that is needed.

Show all your calculations.

Send to tutor

Now get your answers to your study tutor either by post or handing it over personally. You will receive your tutor's assessment and comment back through the post as soon as can be managed.

PART F UNIT REVIEW

1 Return to objectives

Having completed your work on the unit let's review how you feel about each of our four opening objectives. I'll repeat them here and add comment.

You will be better able to:

* USE STATISTICS TO IMPROVE YOUR DECISION-MAKING ABILITY

We've seen how to refine the range and the mean by spotting a rogue figure, make comparisons, check trends and test for a relationship. Because we are 95% certain about any conclusions we make, we are in a better position to make decisions based on those conclusions. For example, suppose an operator appears to be under producing. If, after testing enough values, we find his output level is significantly different from the rest of the operators on the section, we might decide that job training is needed. If you can think of other examples where the use of statistics will help your decision taking, then you have achieved this objective.

You will be better able to:

* INTERPRET STATISTICS AND FORM JUDGEMENTS ON THE INTERPRETATIONS OTHER PEOPLE MAKE

By working through the tests and calculations in this unit we were able to comment upon figures typically met at work. This skill is also useful for checking other people's interpretations of figures. Your ability to do this means you have met this objective.

You will be better able to:

* ASSESS STATISTICALLY THE PERFORMANCE OF YOUR WORK TEAM

The tests and calculations in this unit are tools which can be used to evaluate figures on almost any subject. Practice of many applications will have developed your ability to be confident in the results of your checks. So you have met this objective.

You will be better able to:

* TO USE QUICK AND PRACTICAL STATISTICAL TESTS AT WORK

By knowing which test to apply in a given circumstance and being able to work through the data and calculations carefully, you are able to draw proper conclusions. The conclusions would be more informed than had you guessed - so these tests are of practical value. Your ability across the range of tests means that you have met this objective.

I hope that this work you have done was enjoyable and useful enough to encourage you to continue your studies.

2 Extensions

We recommend that you take up as many of these extensions as you can. They will further increase your understanding and interest. You will also be able to check out a number of the ideas mentioned in the unit. The extra time and effort will prove very worthwhile.

EXTENSION 1

Book: An Introduction to Statistics

Author: A Hall

Publisher: MacMillan Press

EXTENSION 2

Here is an opportunity for you to work through a calculation of the standard deviation.

The STANDARD DEVIATION is a measure of spread which is found from the differences of each individual value from the mean. We'll find the standard deviation for the following set of figures

1, 3, 4, 8, 12, 16, 20, 24

SELF CHECK 1
Time guide
3 mins

We already know how to find the mean of a set of figures. Find the mean by filling in the blanks below.

$\sum$ of values = *_____

Number of values = *_____

Mean is _____ divided by _____, which is _____

The answer to this self check is here in the following words, so you'll need to check out your answers carefully: Sum of values is 88: The number of values is 8. The mean is 88 divided by 8 which is 11.

Now we know our mean, we find the difference of each of our values from the mean. Some of our figures will be higher than our mean of 11. Others will be lower than our mean of 11. We record the difference between each value and the mean.

Let's lay out our figures again, and calculate the differences from the mean by filling in the blanks in the table below.

SELF CHECK 2
Time guide
3 mins

Complete the table below to find the differences from each of our mean of 11. I've put some differences in already:

VALUES	1	3	4	8	12	16	20	24
DIFFERENCES	10	8					9	13

The answer to this self check is here in the following words, so you'll need to check out your answers carefully. Your completed table should look like the one below.

VALUES	1	3	4	8	12	16	20	24
DIFFERENCES	10	8	7	3	1	5	9	13

We now find the standard deviation involved squaring each difference. So if the difference is 10, then squaring this difference gives 100 (10 x 10) - incidentally instead of writing 10 x 10 we could write 10^2 - it means the same thing.

SELF CHECK 3
Time guide
3 mins

Complete this table. Once again I've put some figures in already.

VALUES	1	3	4	12	16	20	24
DIFFERENCES	10	8	7	1	5	9	13
DIFFERENCES SQUARED	100	64				81	169

Here is the completed table:

VALUES	1	3	4	12	16	20	24
DIFFERENCES	10	8	7	1	5	9	13
DIFFERENCES SQUARED	100	64	49	1	25	81	169

We now add up our differences squared:

SELF CHECK 4
Time guide
1 min

Carry out step 4 by summing our differences squared.

The sum of our differences squared is *_____

The answer to this self check is here in the following words so you'll need to check our your answer carefully. The sum of our differences squared is 489.

The standard deviation is the square root of this sum of 489.

SELF CHECK 5
Time guide
2 mins

Find the standard deviation of our set of figures by filling in the blank below.

$\sqrt{489}$ is _____

The answer to this self check is here in the following words, so you'll need to check out your answers carefully. The square root of 489 is slightly over 22. This is easily found using a calculator, by entering the number 489 and then pressing the square root button, which will be marked with the symbol $\sqrt{}$. But we can usually get close enough on our own as we did in the Unit.

EXTENSION 3
page 59

The standard deviation is important to us only in so far as it helps us conduct a number of practical tests. You can explore the standard deviation further by taking up extension 3 if you wish.

Let's summarise the steps we've just taken to find the standard deviation.

STEP 1 Find the mean of our set of figures

STEP 2 Find the difference of each value from the mean.

STEP 3 Square each of our differences.

STEP 4 Sum, or add up, our squared differences.

STEP 5 Take the square root of this sum to find the standard deviation.

EXTENSION 3

Book: Mastering Statistics

Author: T Hannigan

Publisher: MacMillan Press

EXTENSION 4

Book: Rapid Statistical Calculations

Author: M H Quenouille

Publisher: Griffin

These extension opportunities can all be taken up via your Support Centre.

They will either have the materials or will arrange that you have access to the materials. However it may be more convenient to check out the materials with your Personnel or Training people at work - they may well give you access. There are other good reasons for approaching your own people, for example they will become aware of your interest and you can involve them in your development.

3 References

REFERENCE 1	Book:	Supervisory Studies
	Author:	P Betts
	Publisher:	MacDonald and Evans
REFERENCE 2	Book:	People and Work Organisation
	Author:	J G Capey and N R Carr
	Publisher:	Holt
REFERENCE 3	Book:	An Introduction to Statistics
	Author:	A Hall
	Publisher:	MacMillan Press Page 98

These references can all be taken up via your Support Centre. They will either have the materials or will arrange that you have access to the materials. However it may be more convenient to check out the materials with your Personnel or Training People at work - they may well give you access. There are other good reasons for approaching your own people, for example they will become aware of your interest and you can involve them in your development.

If you do wish to make use of these references then don't think that you must read them right through. Just dip into them for a slightly wider treatment of the topic raised by the unit.

I should warn that the language and style of the writing is often difficult to cope with. So don't struggle with it too much. Do what I do when in difficulty, put the book back on the shelf and find a better one!

THE CHOICE SERIES

PRINCIPLES AND PRACTICE OF SUPERVISION

11 Needs and Rewards
12 Enriching Work
13 Workteams
14 Team Leading
15 Leading Change
16 Organisation Systems
17 Supervising in the System
18 Supervising with Authority
19 Workplace Control
20 Taking Decisions

TECHNICAL ASPECTS OF SUPERVISION

30 Looking at Figures
32 Using Graphs
33 Method Study
34 Easy Statistics
35 Quality Control
37 Controlling Output
38 Value Analysis
39 Quality Circles
40 Computers
41 Stores Control
42 Managing Time
43 Descriptive Statistics
45 Supervisors & Marketing

COMMUNICATION

60 Communicating
61 Speaking Skills
62 Writing Skills
63 Communication Systems
64 Orders and Instructions

ECONOMIC AND FINANCIAL ASPECTS

70 Accounting for Money
71 Control via Budgets
72 Cost Reduction
73 Wage Payment Systems
74 The National Economy

INDUSTRIAL RELATIONS

80 Training Plans
81 Training Sessions
82 Discipline and the Law
84 Health and Safety
85 Industrial Relations in Action
86 Equality at Work
87 Hiring People
88 Supervising and the Law